Computers in Biology and Medicine 101 (2018) 1–6

Contents lists available at ScienceDirect

Computers in Biology and Medicine

journal homepage: www.elsevier.com/locate/compbiomed

Novel approach to computerized breath detection in lung function diagnostics

Jaroslav Horáček[a,*], Václav Koucký[b], Milan Hladík[a]

[a] Department of Applied Mathematics, Faculty of Mathematics and Physics, Charles University, Prague, Czech Republic
[b] Department of Paediatrics, 2nd Faculty of Medicine, Charles University, Prague, Czech Republic

ARTICLE INFO

Keywords:
Automated breath detection
Breath end
Lung function testing
Tidal breathing
Multiple breath washout test
Medical algorithm design*MSC:*
92C50
92C55
92-08

ABSTRACT

Background: Breath detection, i.e. its precise delineation in time is a crucial step in lung function data analysis as obtaining any clinically relevant index is based on the proper localization of breath ends. Current threshold or smoothing algorithms suffer from severe inaccuracy in cases of suboptimal data quality. Especially in infants, the precise analysis is of utmost importance. The key objective of our work is to design an algorithm for accurate breath detection in severely distorted data.

Methods: Flow and gas concentration data from multiple breath washout test were the input information. Based on universal physiological characteristics of the respiratory tract we designed an algorithm for breath detection. Its accuracy was tested on severely distorted data from 19 patients with different types of breathing disorders. Its performance was compared to the performance of currently used algorithms and to the breath counts estimated by human experts.

Results: The novel algorithm outperformed the threshold algorithms with respect to their accuracy and had similar performance to human experts. It proved to be a highly robust and efficient approach in severely distorted data. This was demonstrated on patients with different pulmonary disorders.

Conclusion: Our newly proposed algorithm is highly robust and universal. It works accurately even on severely distorted data, where the other tested algorithms failed. It does not require any pre-set thresholds or other patient-specific inputs. Consequently, it may be used with a broad spectrum of patients. It has the potential to replace current approaches to the breath detection in pulmonary function diagnostics.

1. Introduction

Breath detection (i.e. finding the spot where expiration ends and the consecutive inspiration starts) is a crucial step in pulmonary function testing (PFT). It is a starting point for the computing of various clinically significant indices, performing regression analyses or making predictions. With the increasing importance of PFT as a diagnostic tool, new methods of PFT and approaches to data analysis are required especially in infants and toddlers (i.e., uncooperative children). In this age category, precise raw data analysis is of utmost importance, as the infant PFT is very prone to technical errors. Based on our clinical experience, the current PFT algorithms suffer from severe inaccuracy, which may lead to difficult and time-consuming interpretation of results or even raw data rejection.

Although breath detection is a relatively easy task for a physician, the automated detection by a computer remains a challenge, especially in cases of severely distorted data (e.g., as a result of young patients not cooperating well, severe drift etc.). An approach to the breath detection analysis is primarily determined by the signals being measured. Usually, a time-flow signal is captured. In this situation, there exist two basic algorithms for breath detection, that have already been proposed – threshold and smoothing approach, each with numerous modifications and extensions in an attempt to achieve greater reliability and accuracy [1]. The threshold approach rejects any breath having parameters below pre-set threshold. On the other hand, the smoothing approach smooths the signal to eliminate spurious breath endings. Despite the significant progress done in this field, clinicians are still facing situations in which the measured signal is too distorted to be automatically analysed.

Multiple breath washout test (MBW) is an example of a highly sensitive method recently introduced into clinical practice [2] [3], or [4]. It offers an important insight into early stages of several chronic lung diseases [5], [6]. Moreover, it does not require active breath manoeuvres and can be performed on infants during tidal breathing.

* Corresponding author.
E-mail addresses: horacek@kam.mff.cuni.cz (J. Horáček), vaclav.koucky@seznam.cz (V. Koucký), hladik@kam.mff.cuni.cz (M. Hladík).

https://doi.org/10.1016/j.compbiomed.2018.07.017
Received 5 March 2018; Received in revised form 28 July 2018; Accepted 28 July 2018

Consequently, this method may yield clinically extremely relevant information. However, it requires a special approach to raw data analysis which the current algorithms may not offer. In comparison with the conventional methods which are based solely on flow, volume and pressure measurements and estimate primarily airway resistance (e.g. bodypletysmography, tidal breath analysis etc.), MBW brings a new dimension to raw data – the gas concentration signal (O_2, CO_2, inert gas). A current commercial software (Spiroware, Ecomedics, Duernten, Switzerland) uses concentrations only for constructing washout curves. However, this information may be also used for breath detection. The aim of our study was to design and justify a new and robust algorithm for breath detection using not only time-flow data but also the gas concentration signal. Such a breath detection algorithm can significantly outperform the current threshold-based algorithms. Moreover, its key ideas have the potential to contribute to the general design of the medical algorithms.

2. Materials and methods

Raw data from nitrogen multiple breath washout test were used as an input for the breath end analysis. Data were captured by the machine Exhalyzer D, Ecomedics, Duernten, Switzerland with software Spiroware 3.2.0, following the relevant recommendations by European cystic fibrosis society (ECFS) [7] and European Respiratory Society (ERS) [8]. The raw data were stored in. txt files with a specific structure containing time, flow, oxygen (O_2), carbon dioxide (CO_2) and molar mass (MM) signals, all measured every 5 m s (sampling frequency of 200 Hz).

2.1. Our algorithm

Our algorithm (Alg-OUR) was programmed in the free software GNU Octave, version 4.0.0. and works in several steps, which are outlined below. A depiction of each step can be found in Fig. 1.

1. Function load – flow and CO_2 concentration data in time are stored in working memory and visualised.
2. Zero-crossings detection – a zero-crossing is defined as a time spot, where the air flow changes its direction from minus to plus (see the comment on general physiology of respiratory tract in section Discussion). All the zero-crossings in flow raw data are detected and numbered from 1 to N, where N is the total number of zero-crossings. They form a set of potential breath ends.
3. For each −/+ zero-crossing at time T, the nearest peak of CO_2 curve (i.e. local maximum) is found and attributed to the time T.
4. The volume of each inhalation and exhalation (V_{in}, V_{out}) corresponding to the time T is calculated by integration of the flow curve (using simple trapezoidal rule).
5. The zero-crossings with corresponding CO_2 peaks of insufficient concentration (i.e., less than 2% - see comment in section Discussion) are discarded; the numbering of zero-crossings is preserved. Next, our goal is to discard zero-crossings that do not form a breath end or capture intervals of zero-crossings that belong to the same breath. Initially, we view each zero-crossing to be a singleton interval ($[b, b]$). Next, the algorithm is going to discard or merge some of these intervals (steps 6 to 8).
6. Two intervals of zero-crossings $[a, b]$ and $[c, d]$ are merged if the CO_2 concentration between b and c does not drop below 0.5%. Consequently, a new interval $[a, d]$ instead of the previous two is created. This process is repeated until there exists no such a pair of intervals. Note that in an interval $[a, b]$, a can be equal to b.
7. The two consecutive intervals of zero-crossings $[a, b]$ and $[c, d]$ where $c = b + 1$ are merged if the ratio of volumes V_{in}/V_{out} for zero-crossing b is greater than 5 (see the comment in section Discussion). This process is repeated until there exist no such a pair of intervals.
8. The upper bounds of the remaining intervals (even tight ones - $[a, a]$) are marked as the breath ends (i.e., from $[a, b]$, it is b, from $[a, a]$, it is a).

NOTE. The order of the steps 5, 6 and 7 can not be changed; otherwise the algorithm produces incorrect results.

For the sake of comparison, the most commonly used flow threshold algorithm (originally described in Ref. [9]) was implemented in our software. Two different thresholds (10 ml for Alg3-0.01 and 250 ml for Alg3-0.25) according to the age of the patient and an additional plausibility check were used as specified in Refs. [1] [9], and [10].

2.2. Data characteristics

To test the clinical usefulness and accuracy of our newly developed algorithm, we compared it with representatives of the currently used algorithms on real patient data. We intentionally selected severely distorted measurements, which are, in our experience, very difficult to be automatically analysed by the current software. In total, 47 anonymized traces (A-files) obtained from the 19 patients were enrolled on the trial. Although all the patients were primarily measured for clinical purposes (diagnostics, follow up, etc.), their legal representatives gave informed written consent to the raw data analysis for research purposes. Such an approach was in general approved by the local ethics committee. The patients' characteristics are stated in Table 1. The rationale for the intentional selection of severely distorted data was, that only such data offer the possibility to test the performance of breath detection algorithm properly. The analysis of regular breathing is no challenge for current breath detection approaches anymore.

2.3. Comparison of algorithms

The raw data were analysed in four different ways:

1. an analysis performed by our algorithm described above (Alg-OUR),
2. an analysis performed by the previously described algorithms (Alg3-0,01 and Alg3-0,25) that are implemented in our software,
3. an analysis performed by the commercial package Spiroware 3.2.0 (Alg-Spi),
4. a manual analysis performed by two specialists experienced in PFT.

After loading the respective A-file into our software, the number of breaths detected by Alg-OUR, Alg3-0,01 and Alg3-0,25 were calculated. The A-file was also loaded into Spiroware and the number of breaths was estimated using the functionality of this commercial software. Afterwards, two PFT specialists inspected the data from each A-file independently. The inspection was done in the interface of our software. It enables visualization of flow, volume and CO_2 concentration, while at the same time visualisation of breath ends found by the respective algorithms. Such visualization enables both the estimation of the number of true breaths (reference number of breaths RNB) and simultaneously the localization of falsely positive/negative breaths as analysed by different algorithms.

3. Results

All the A-files included in our testing could be successfully analysed by all the implemented algorithms. The analysis time was longer for Alg-OUR than for the threshold algorithms (1.35 ± 0.23s vs. 0.12 ± 0.01s, $p < 0.001$). The manual analysis took much longer; the average analysis time was roughly estimated to be between 100 and 180s.

The two specialists in PFT working independently detected the same number of breaths in 35 out of 47 A-files (74%). In the remaining cases, differences were not larger than two breaths. These cases were re-analysed by the two specialists jointly in order to reach consensus and reference number of breaths (RNB) was assigned to each A-file. Finally, 2861 true breaths in 47 A files were included.